being a question that was identified in the corpus that matches the question posed by the user. At step **720**, the process retrieves a set of follow-up postings that pertain to the selected question with the retrieval being from the harvested discussions stored in harvested discussions data store **320**. In addition, real-time retrieval of follow-up postings can be performed by collecting such follow-up postings from web sites **300**. For example, if an answer to a question was recently provided to the online threaded discussion but such answer has not yet been harvested by the process shown in FIG. **4**, then these up-to-date postings can also be gathered in addition to the posts previously gathered by the discussion harvesting routine. The follow-up postings are stored in follow-up postings data store **725**. At step **730**, the rhetorical structure and informational structure of the online discussion thread are analyzed.

[0048] At step **740**, the first follow-up posting is selected. At step **750**, the selected follow-up posting is analyzed according to its conversational move (such as 'answer', 'clarify', 'reject') using existing dialogue parsing techniques. At step **760**, the selected follow-up post is labeled based on the conversational move corresponding to the follow-up post. This results in a tree of contributions and responses, as well as a set of candidate answers for the posed question. This tree of contributions is stored in contribution tree data store **770**. A determination is made as to whether there are additional follow-up postings to process (decision **775**). If there are additional follow-up postings to process, then decision **775** branches to the "yes" branch which loops back to select and process the next follow-up post as described above. This looping continues until all of the follow-up posts have been processed, at which point decision **775** branches to the "no" branch.

[0049] At predefined process **780**, the contribution tree is pruned and candidate answers are identified and stored in data store **780** (see FIG. **8** and corresponding text for further processing details). A determination is made as to whether there are additional questions from closest matching question posts data store **360** that need to be processed (decision **790**). If there are additional questions to process, then decision **790** branches to the "yes" branch which loops back to select the next question from closest matching question posts data store **360**. This looping continues until all of the questions that were found to match the question received from the user have been processed, at which point decision **790** branches to the "no" branch and processing returns to the calling routine (see FIG. **6**) at **795**.

[0050] FIG. **8** is a depiction of a flowchart showing the logic performed by the system to prune a contribution tree of unneeded or superfluous contribution posts. Processing commences at **800** whereupon, at step **810**, the first labeled contribution is selected from contribution tree data store **770**. At step **820**, the process analyzes the selected contribution for possibly pruning the selected contribution from the contribution tree. This analysis includes an analysis as to whether the contribution leads to a new (different) question, whether the contribution is overly deep in terms of plys, as well as any other additional pruning criteria that may wish to be employed by the question answering system. A determination is made, based on the analysis, as to whether the selected contribution should be kept in the contribution tree (decision **825**). If the analysis reveals that the contribution should not be kept in the contribution tree, then decision **825** branches to the "no" branch whereupon, at step **380**, the process deletes the selected contribution from contribution tree data store **770** with the deletion also deleting any other contributions in the contribution tree that stem from the selected contribution. On the other hand, if the analysis reveals that the selected contribution should be kept in the contribution tree, then decision **825** branches to the "yes" branch bypassing step **830**.

[0051] A determination is made as to whether there are additional contributions in the contribution tree to select and analyze for possible pruning (decision **840**). If there are more contributions to process, then decision **840** branches to the "yes" branch which loops back to select and process the next contribution as described above. This looping continues until all of the contributions have been processed, at which point decision **840** branches to the "no" branch to perform candidate answer identification.

[0052] At step **850**, the process selects the first contribution labeled as an 'answer' that remains in contribution tree **770** after the pruning operation. At step **860**, the process evaluates the type of the selected answer, such as whether the answer is answering a "how?" question, a "how much?" question, a "how many?" question, a "who?" question, a "where?" question, a "why?" question, a "when?" question, a "to what extent?" question, and the like. A determination is made as to whether the selected answer contribution has an answer type that matches the type of answer required by the user's question (decision **870**). For example, the user may be asking "how" to change the oil on a particular automobile, but the selected answer may be answering "why" it is important to frequently change the oil on this particular automobile. If the selected answer's type matches the type of answer required to answer the user's question, then decision **870** branches to the "yes" branch whereupon, at predefined process **880**, the selected answer is scored and included as a candidate answer and stored in scored answers data store **370** (see FIG. **9** and corresponding text for further processing details). On the other hand, if the selected answer's type does not match the type of answer required to answer the user's question, then decision **870** branches to the "no" branch bypassing predefined process **880** and the selected answer is not included in the set of answers stored in scored answers data store **370**.

[0053] A determination is made as to whether there are more remaining contributions to process to identify potential candidate answers (decision **890**). If there are more remaining contributions to process, then decision **890** branches to the "yes" branch which loops back to select the next contribution labeled as an 'answer' from contribution tree **770** and evaluate and process the answer as discussed above. This looping continues until all of the remaining contributions have been processed, at which point decision **890** branches to the "no" branch and processing returns to the calling routine (see FIG. **6**) at **895**.

[0054] FIG. **9** is a depiction of a flowchart showing the logic performed to add candidate answers to the set of answers identified for consideration as the most likely answer to the identified question. Processing commences at **900** whereupon, at step **910**, evidence for the answer that is being scored is aggregated and merged from various online discussions. Evidence is retrieved from the harvested discussion data that has been stored in harvested discussions data store **320**. In one embodiment, evidence can be gathered from other posts that use different phrasing or wording for their answers as well as from posts gathered from